STAT 453/553: Biostatistics and Bioinformatics Syllabus
Spring 2013
TTH 4:15-5:30pm, 286 BRC

Overview

This course provides an overview of statistical methodology useful in the practice of modern biostatistics. There are three major topics covered in the course, each an introductory level for basic data analysis in these areas: biostatistics for epidemiology, clinical trial design, and bioinformatics. More specific topics are listed in the attached daily syllabus. Previous versions of this course included epidemiology, survival analysis, and clinical trials. However, because of the wide availability of genetic data (much of it public), many "biostatistics" research projects now include some aspect of genetic data or information; this includes survival analysis studies and clinical trials. Therefore, it is important that students have some exposure to the basic statistical approaches to analyzing genetic data. In this course, the student will learn about different types of genetic data that are of current interest and basic approaches to their statistical analysis. Real biomedical applications serve as context for evaluating assumptions of statistical methods and models. S-Plus (R) serves as the preferred computing software, but students may use statistical software of their choice. However, the instructors may not be able to provide support for certain software.

Instructors: There are two instructors for the course. Each instructor will cover about 7 weeks of the course. Dr. Yuan will cover biostatistics for epidemiology and clinical trials and Dr. Li Zhang will cover statistical genetics and bioinformatics. All classes will be held in room 285 of Biosciences Research Collaborative (BRC) Building at Rice University.

Instructor:	Ying Yuan	Li Zhang
Email:	yyuan@mdanderson.org	lzhangli@mdanderson.org
Voice:	713.563.4271	713-563-4298
Location:	FCT4.6084 @ MDA Pickens Towers	FCT4.5034@ MDA Pickens Towers
OfcHrs:	By Appt	By Appt

Textbooks:

1. **Statistics for Epidemiology**, by Nicholas Jewell. Publisher: Chapman & Hall/CRC

2. **Bioinformatics-The machine learning approach**, by Pierre Baldi and Soren Brunak. Publisher: The MIT Press.

Homework: Each of the two major parts will include 2-3 assignments, at least one of which will be data based. All students are required to complete the assignments. Homework will be submitted at the beginning of class on the due date. If circumstances beyond the student's control arise and an assignment cannot be submitted on the due date, an instructor should be contacted prior to the due date. With an instructor's permission, late homework may be accepted within one week of the due date. All decisions will be made on an individual student basis and the final decision rests with the instructor assigning the homework. A penalty of 10 percentage points will be applied to late homework.

Website: http://odin.mdacc.tmc.edu/~yyuan/ has more information about the Biostatistics part of the course and datasets for the homework.

Examinations: There will be one in-class exam for each of the two parts in the course. No final examination will be given. All students are required to take both exams.

Course Grade and STAT 453 vs 553

The material covered in the classroom serves two different course, STAT 453 (undergraduate) and 553 (graduate). Although the lectures will cover the same material for both courses, those enrolled in STAT 553 will be required to answer either additional questions or more challenging questions on the both the homework assignments and in-class exam. Although the requirements for the two courses are the same, different grading scales will be applied to the two courses according to their relative level. Both courses will be graded on the following component basis.

Component	%
Biostat Hwk	20
Biostat Exam	30
Bioinf Hwk	20
Bioinf Exam	30

Policy on Submitted Work

Students are permitted to work together on their homework assignments unless otherwise noted. However, each student is expected to submit their own individual assignment based on their own individual expression. Data based projects will be expected to be submitted as reports written with a word processor; page limits will be given.

Disability Statement

Any student with a disability requiring accommodations in this course is encouraged to contact me after class or during office hours. Additionally, students will also need to contact Disability Support Services in the Ley Student Center.

TENTATIVE DAILY SYLLABUS

Week	Date	Topic	Readings*	Notes	Instructor
1	1-8-T	Introduction		Overview the course	Yuan
	1-10-Th	Study Design	SE: 3, 5	Cohort, case-control and observational studies	Yuan
2	1-15-T	Disease-Exposure Association	SE: 4	Risk, odds, odds ratio, relative risk, standard errors	Yuan
	1-17-Th	Contingency Tables: Association	SE: 6	Chi-square test	Yuan
3	1-22-T	Contingency Tables: Confounding	SE: 9	Mantel-Haenszel	Yuan
	1-24-Th	Contingency Tables: Interaction	SE: 10	Test of homogeneity	Yuan
4	1-29-T	Logistic Regression: introduction	SE: 12	Including review of maximum likelihood estimation	Yuan
	1-31-Th	Logistic Regression: estimation	SE: 13	Interaction	Yuan
5	2-5-T	Matched studies	SE: 16		Yuan
	2-7-Th	Matched studies	SE: 16		Yuan

6	2-12-T	Case study I: Risk assessment models for cancer	HO		Wang
	2-14-Th	Case study II: Gaussian mixture models for genotyping using microarray data	HO		Wang
7	2-19-T	**Study days**			Yuan
	2-21-Th	**Midterm Exam**		**In class exam: Biostatistics**	Yuan
8	2-26-T	**Spring Break**			
	2-28-Th	**Spring Break**			
9	3-5-T	Introduction to topics in bioinformatics			Zhang
	3-7-Th	High throughput technologies in genomics			Zhang
10	3-12-T	Common tools used in bioinformatics; Introducing R			Zhang
	3-14-Th	Shannon Entropy and mutual information			Zhang
11	3-19-T	Non-linear relationship; Maximum information coefficient			Zhang
	3-21-Th	**Midterm Recess**		No class	
12	3-26-T	Mutual information used in network analysis			Zhang
	3-28-Th	Hidden Markov Models			Zhang
13	4-2-T	HMM applied in sequence analysis			Zhang
	4-4-Th	Network analysis-1			Zhang
14	4-9-T	Network analysis-2			Zhang
	4-11-Th	Network analysis-3			Zhang
15	4-16-T	Study days			Zhang
	4-18-Th	**In class exam of bioinformatics**		**Last Day of Classes**	Zhang

*SE: Statistics for Epidemiology textbook; HO: Handout